②

DTIC FILE COPY

AD-A218 915

The Artificial Intelligence and Psychology Project

DTIC
ELECTE
MAR 12 1990
S E D

Departments of
Computer Science and Psychology
Carnegie Mellon University

Learning Research and Development Center
University of Pittsburgh

90 03 12 066

2

AHA!: A CONNECTIONIST PERSPECTIVE ON PROBLEM SOLVING

Technical Report AIP - 38

Craig A. Kaplan

Carnegie Mellon University
Department of Psychology
Pittsburgh, PA. 15213

8 June 1988

This research was supported by the Computer Sciences Division, Office of Naval Research and DARPA under Contract Number N00014-86-K-0678. Reproduction in whole or in part is permitted for purposes of the United States Government. Approved for public release; distribution unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) AIP - 38	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Carnegie-Mellon University 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Computer Sciences Division Office of Naval Research
6c ADDRESS (City, State, and ZIP Code) Department of Psychology Pittsburgh, Pennsylvania 15213	7b ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION Same as Monitoring Organization 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0678
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS p4000ub201/7-4-86

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
N/A	N/A	N/A	N/A

11 TITLE (Include Security Classification)

AHA!: A connectionist perspective on problem solving

12 PERSONAL AUTHOR(S)
Craig A. Kaplan

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	1988 June 8	7

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Connectionist models, Problem solving, Insight

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The AHA! model is proposed as a demonstration of what connectionism might have to offer the study of problem solving. Bridging the Gestalt and Problem Space theories of problem solving, AHA! simulates serial search at a macro-level while incorporating (at a micro-level) the Gestaltist idea of a dynamic interaction between parts of the problem, and the goals and knowledge of the problem solver. AHA! exhibits a number of problem solving phenomena including insight, directed search, goal fixedness, einstellung, functional fixedness, and responsiveness to the salience of problem features. It provides not only qualitative fits to human data from functional fixedness experiments, but also a framework for exploring the potential outcomes of variations which could be carried out in future experiments.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Alan L. Meyrowitz	22b TELEPHONE (Include Area Code) (202) 696-4302 22c OFFICE SYMBOL N00014

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

AHA!: A Connectionist Perspective On Problem Solving

Craig A. Kaplan

Carnegie-Mellon University

What can connectionism offer Cognitive Scientists interested in problem solving? How might Connectionist ideas fit with existing theories of problem solving -- specifically the Gestalt approach and the problem space approach advocated by Newell & Simon (1972)? Finally, can a Connectionist model of problem solving **perform**? Can it easily account for a variety of problem solving phenomena?

The Gestalt theory of problem solving is a good place to begin answering these questions. One of the great appeals of the Gestaltist perspective on problem solving is that the problem solver and all of the elements of the problem are considered to be dynamically inter-related The solution is a gestalt that emerges from the simultaneous influence of each part of the problem. During a flash of insight or "restructuring," all of the parts suddenly fit together in a new way forming a solution. Analyzing this solution gestalt into its component parts seems to miss the very essence of the Gestaltist view of problem solving, namely the dynamic relations between the parts. Yet without a more rigorous analysis, the Gestaltist theory remains inescapably vague and subject to the liberties of individual interpretation.

Newell and Simon's conception of problem solving as search through a problem space provides rigor. The success of that rigor is well known and ranges from AI programs capable of making scientific discoveries, to accounts of human behavior in a wide variety of problem domains, including puzzle problems requiring "insight" (Kaplan & Simon work in progress). Some key elements of the theory are: 1)Problem solving behavior can be described as search through a state space with each state corresponding to a configuration of the problem and/or a specification of the knowledge state of the problem solver. 2)Movement from one state to another state is accomplished by applying an operator corresponding to performing some physical or mental action. 3)Heuristics, or rules of thumb, can help select which of a number of potential operators to apply.

Typically, the search perspective has been used to describe problem solving behavior occurring on a macro-level time scale of seconds as opposed to the more detailed micro-level of processing where states and operators are actually generated. It is here, at this micro-level of operator or idea generation, that Gestalt ideas might prove more fruitful if they could be formalized. Connectionism offers the possibility of rigorously modeling the dynamic interactions described by the Gestaltists at a micro-level, while retaining compatibility with the notion of serial search through a problem space at a macro-level. The most direct way to explore this possibility is via a connectionist network model incorporating an interactive activation mechanism.

Consider a problem solving model consisting of inter-connected units corresponding to the features of a problem and to the goals and knowledge of the problem solver. Through a process of interactive activation (and inhibition) these units "settle" into a stable state -- a process that roughly corresponds to generating an operator, applying it, and arriving in a new state in Newell and Simon's problem space model. For the sake of interpretability, and as a convenient means of providing feedback, the network model could explicitly represent possible states in the problem space by a set of units. After a number of processing cycles, the most active "state" unit could receive positive or negative external input, corresponding to a favorable or unfavorable evaluation of the idea it represents. This input would spread through the unit's connections, either reinforcing the state or causing a new state to be generated. The model's macro-level behavior would appear as a sequential generation of states, or movement through a problem space. However the Gestaltist ideas of dynamic interaction and simultaneous influence would be at work at the micro-level, generating each successive state.

AHA! as a Problem Solver

The AHA! (Associative Hierarchical Activation) model is a concrete instantiation of the ideas just mentioned, in a connectionist network. AHA! was designed to simulate solving variations of Duncker's Box problem (Duncker 1945). Briefly, the Box problem consists of trying to attach a candle to the wall (so that it can burn normally) using a candle, matches, a box of tacks. The insightful solution consists of tacking the box to the wall *and* placing the candle in it. Subjects find it easier to "see" this solution when the tacks have been first emptied from the box (the tacks-not-in -box condition) as compared to a condition where the box contains the tacks (the tacks-in-box condition). For the sake of simplicity, AHA! tries to achieve only a partial solution to the Box problem. AHA!'s criteria for a successful solution is generating either the idea *attach the box with tacks* or the idea *support the candle with box*. Research with other insight problems (Kaplan 1986) indicates that a human subject capable of generating either of these partial solutions ought to find the remaining steps in the the total solution trivial.

Figure 1 depicts the four layered organization of AHA!. Units in the PI (Perceived Item) layer correspond to actual items in the physical problem environment such as the candle and matches. The box of tacks has two complementary representations at this level: box-of-tacks, or box and tacks. The RA (Role Assignment) layer contains units representing the assignment of items to the roles of instrument and object (e.g. a candle-as-object unit, abbreviated **c-obj**). Units in the topmost F level correspond to functions that the items might possess, or equivalently, actions that might be taken with them. (e.g. **attach**, or **ignite**). Possible states in AHA!'s problem space are represented by triples at the T level. Each triple consists of a function (action) relating an item in the role of object to an item in the role of instrument (e.g. ignite candle-as-object [with] matches-as-instrument, abbreviated **i-c-m**). Connections between units in adjacent layers are excitatory, while the connections within a level are generally inhibitory. Activation spreads bi-directionally (i.e. interactively) between units as the network settles. The T unit with the highest activation after any given settling is considered to be the state generated by the model.

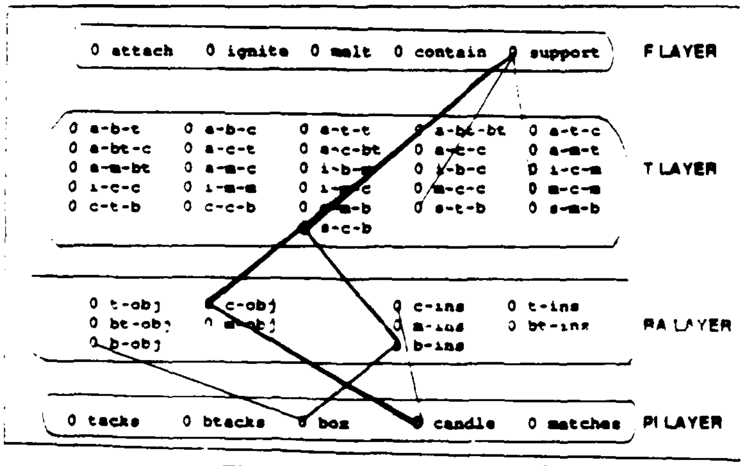

Figure 1: The AHA! Model

Before looking at the detailed specifications of the model, consider these problem solving phenomena that follow from the general characteristics of the architecture:

Insight & Restructuring

AHA! has no special "insight mechanism." AHA! accounts for insight in the same way it accounts for the generation of any new problem solving idea. A conspiracy of problem (and problem solver) elements, whose own activations are inter-dependent and responsive to simple feedback from the environment, activates one T unit more than its peers. As discussed below, a sudden change of representation, or restructuring, often accompanies AHA!'s insight -- not through any special mechanism, but again as a consequence of the model's interactive processing characteristics.

Directed Search

Consider AHA!'s generation of the idea/state **m-c-m** (melt candle with matches). Because this state is not a partial solution by the model's criteria, it would receive negative feedback. However, while it was active, it would have had a chance to activate related concepts (e.g. the objects **candle** and **matches**). Hence the next generated state, perhaps **i-c-m** (ignite candle with matches), is likely to contain elements in common with the previous state. Using what amounts to semantic priming, AHA! is able to exhibit the phenomena of directed search. Like a human problem solver, AHA! will tend to generate bursts of related ideas, exploring the problem space systematically rather than generating states haphazardly.

Goal Fixedness

Goals correspond to a pattern of activation over a number of units. For example, external input applied to the units **attach** and **c-obj**(candle as object) corresponds to the goal: *attach the candle.* In most AHA! simulations this initial input is allowed to decay. Without constant input from outside the network, the model's goals fluctuate with the activations of other related units in the network, thus allowing AHA! to change goals flexibly during the course of solving the problem -- just as human subjects typically do. However, clamping these "goal inputs" causes AHA! to adhere more rigidly to the initial goal of the problem. Just as overly goal-directed human subjects often persist in using the same approach despite repeated failures, the Goal Fixed AHA! model takes longer to generate a partial solution, needlessly revisiting a state that had been previously generated and rejected.

Einstellung

Suppose that instead of clamping external input to units corresponding to a goal, we provide initial input to one of the partial solution states represented at the T layer. This primed solution state might correspond to the problem solving set, or Einstellung, evoked in human subjects who have repeatedly solved a number of problems in the same way. Like human subjects suffering from Einstellung, AHA! tends to choose the "primed" solution when solving a problem that has multiple solutions (see below).

Functional Fixedness & Salience Effects.

The PI level in AHA! includes alternative ways of representing the box and the tacks in the Box problem, corresponding to Duncker's belief that subjects perceive the (function of) the box differently in the two conditions of his experiment. Providing input to one of these perceptions and not the other corresponds to increasing the salience of the former. If the tacks-in-box perception (**btacks**) is favored over the box-separate-from-tacks perception, then AHA! exhibits the phenomenon of functional fixedness (i.e. it experiences difficulty in generating a partial solution). As will be seen, AHA! provides qualitative fits to the data from two separate experiments involving the Box problem, and is consistent with both Functional Fixedness (Duncker 1945) and salience (Glucksberg & Weisberg 1966) interpretations of the Box problem's difficulty.

Detailed Specifications of AHA!

Interactive Activation

AHA! uses the interactive activation and competition (IAC) architecture (McClelland 1981, McClelland & Rumelhart 1988). The network consists of processing units that are organized into layers and interconnected using bi-directional excitatory and inhibitory connections. The activation of a particular unit depends upon both its current activation and the net input to the unit from other units and from outside the network. The net input to a unit i is calculated by:

$$input_i = S(Ext_i) + E\sum_j e_{ij} - I\sum_j i_{ij}$$

Ext_i corresponds to any input to unit i from outside the network, e_{ij} are the activations of units with excitatory connections to unit i, and i_{ij} are the activations of units with inhibitory connections to unit i. The constants S, E and I (all set to .05 for the simulations described below) scale the strength of the external input, the excitatory input from other units, and the inhibitory effect of other units respectively. The actual change in the activation of a unit i must take into account not only the net input, but also unit i's current activation, a_i. If the net input is excitatory (i.e. >= 0) then :

$$\Delta a_i = (max - a_i)input_i - decay(a_i - rest).$$

If the net input is inhibitory (i.e. < 0) then:

$$\Delta a_i = (a_i - min)input_i V - decay(a_i - rest).$$

The parameters *max* and *min* correspond to the maximum (+1) and minimum (-.2) activations that a unit can assume, while *decay* (.05) specifies the rate at which a unit has the tendency to decay back to its resting value, *rest* (-.1).

Interconnectedness

Connections between layers of units are excitatory, with units being connected to conceptually related units in other layers. Some of these connections are illustrated in Figure 1. All units within a layer have mutually inhibitory connections with two exceptions: 1) At the RA level, there are no inhibitory connections between items used as objects and different items used as instruments, since such connections would be nonsensical. 2) At the PI level, only the two representations that Duncker hypothesized as alternatives are put into competition (i.e. **btacks** inhibits the mutually excitatory pair of **box** and **tacks**).

The units themselves were derived from Duncker's description of the problem, pilot data from subjects asked to solve the problem, and (in the case of the T units) by constructing all triples that were physically possible given real world materials (e.g. real tacks, a real box, etc.)

Feedback

A single update of the activation values of all the units in the the network according to the rules described above constitutes one cycle. AHA! was allowed 100 cycles to settle each time a change was made to the external input. After settling, the author observed the activation values of the T units, and considered the most active unit to be the next generated state. In the event of a tie of activation values, one of the tied units was selected arbitrarily.[1] If the state did not correspond to a partial solution of the

[1] A subsequent analysis revealed that the qualitative fits to the data and general trends are not affected by the way in which ties are decided.

box problem (i.e. was not **a-b-t** or **s-c-b**), then the T unit was clamped with a -.5 external input, and AHA! was allowed to settle for another 100 cycles.

Simulations

Five simulations are discussed in this paper. The first three correspond to conditions of Functional Fixedness experiments performed by Duncker and Glucksberg & Weisberg, while the remaining two illustrate the phenomena of Goal Fixedness and Einstellung. In each simulation, a high external input (.9) was applied to the units **attach** and **c-obj** for the first 100 cycles, corresponding to the initial problem solving goal of *attach the candle*. A constant input (.5) was also applied to the units **matches** and **candle**, corresponding to the idea that these items remained moderately salient in all of the variations of the problem. The differences between the simulations will be pointed out as we proceed.

Simulations of Functional Fixedness & Salience Effects

To make sense of the first three simulations, we need to understand their correspondence to experimental conditions involving human problem solvers. While Duncker had two conditions in his experiment, one in which the tacks were contained in a box, and one in which the tacks were separate from the box, Glucksberg & Weisberg presented subjects with three versions of the Box problem. In the All Label version, subjects saw a picture of a box containing tacks, a candle, and some matches. The words BOX, TACKS, CANDLE, and MATCHES were also present in the picture, with arrows pointing from labels to items. In the No Label version, the same picture was presented without the labels. In the Tacks Label condition, the same picture was again presented, but this time there was a single label TACKS with an arrow pointing to the box of tacks. The task, as in Duncker's experiment, was to find a way to attach the candle to the wall using the items shown in the picture.

Differences between the All Label, No Label, and Tacks Label conditions were simulated by providing different inputs to the units **btacks**, **box**, and **tacks**. In the All Label simulation, both **box** and **tacks** were clamped with moderate input (.5) corresponding to the fact that the labels made each of the items salient. **B-tacks**, the unit corresponding to a perception of the box of tacks as a single unit whose function was the same as that of tacks, received no input. In the No Label simulation, **btacks** was clamped with a moderate input (.5) while **box** and **tacks** received no input. Finally, in the Tacks Label simulation, **btacks** was clamped with a high (.9) input, corresponding to a particularly strong tendency to view the box of tacks as a single item possessing only the properties of tacks. Again, **box** and **tacks** received no input in the Tacks simulation.

Note that the All Label condition roughly corresponds to Duncker's tacks-separate-from-box condition, since the both the label BOX and physically separating the box from the tacks can be interpreted as making the box more salient. Similarly, the No Label condition corresponds to Duncker's tacks-in-box condition, since subjects tend not to consider the box in either of these conditions. Due to this rough equivalence between the two experiments, the same AHA! simulations have been used to model the first two conditions in both experiments. Since Duncker has no condition corresponding to the Tacks Label condition of Glucksberg & Weisberg, the Tacks Label simulation fits data from a single experiment only.

Fits to the Human Data. As is readily apparent from Table 1, AHA! generates the least number of states in the All Labeled condition (4), considerably more states in the No Label condition (10), and even more in the Tacks Label condition (11 before it is unable to generate further states because all the

Table 1: State Sequences Generated by Five AHA! Simulations

All Labeled	No Labels	Tacks Label	Goal Fixed	Einstellung
a-c-t	a-c-bt	a-c-bt	a-c-t	a-c-t
a-c-bt	a-c-t	a-c-t	a-c-bt	s-c-b*
a-c-c	a-c-c	a-c-c	a-c-c	
a-b-t*	i-c-c	a-m-bt	a-c-t	
	m-c-c	a-bt-bt	a-b-t*	
	m-c-m	a-m-c		
	i-c-m	i-m-c		
	i-b-m*	a-m-t		
	i-m-m	i-m-m		
	s-c-b	c-m-b*	* indicates the first state	
		s-m-b	to make use of the box.	
		STUCK		

Table 2: Data From Human Subjects Solving the Box Problem

EXPERIMENTAL CONDITION	% SUBJECTS WHO SOLVED PROBLEM	AVERAGE # OF PRE-SOLUTIONS	% WHOSE FIRST SOLUTION USES BOX
Tacks not in Box (Duncker)	100%	1.3	not reported
Tacks in Box (Duncker)	42.9%	2.3	not reported
All Labeled (G & W)	100%	not reported	95%
No Labels (G & W)	85%	not reported	65%
Tacks Label (G & W)	77.1%	not reported	54.3%

activations at the T layer have become negative). From these data, we would conclude that the conditions are progressively more difficult and we might expect progressively fewer subjects to solve the problem in the more difficult conditions. Furthermore we might expect those subjects who did solve the problem in the more difficult conditions to require more solution attempts, just as AHA! does.

Data from human subjects, presented in Table 2, confirms our expectations based on the model's behavior. Both human subjects and AHA! find the same variations of the Box problem difficult, and make more solution attempts on the more difficult variations. The Glucksberg & Weisberg result that fewer subjects use the box in their first proposed solution as the problem conditions get more difficult, is also reflected in AHA!'s behavior. The asterisks in Table 1, marking AHA!'s first proposal of using the box in a solution, occur progressively later in the simulation traces corresponding to the more difficult variations of the Box problem. 1

Simulations of Goal Fixedness, Einstellung, & Restructuring

Goal Fixedness. To simulate the lack of flexibility exhibited by overly goal-directed problem solvers, the All Labeled simulation was modified. Instead of removing the external inputs to **attach** and **c-obj** (corresponding to the goal of *attach the candle*) after 100 cycles as was done with all the other simulations, these inputs were left "clamped on." As Table 1 indicates, the result was a delay in generating the partial solution **a-b-t** (attach box with candle). A close look at the sequence of states reveals that the delay is caused by a revisitation of the state **a-c-t** (attach candle with tacks). Although this state was rejected earlier, it received such strong support from the clamped goal that it was regenerated.

Einstellung The All Labeled simulation was again modified to simulate Einstellung This time the partial solution state **s-c-b**(support candle with box) was primed with a high external input for 100 cycles Table 1 shows that whereas the All Labeled simulation ends up generating **a-b-t**(attach box with tacks), the Goal Fixed simulation generates **s-c-b**. Thus a complete change in the direction of problem solving resulted from priming one of the partial solution states

Restructuring Interestingly, in the Einstellung simulation, AHA! shifts goals from *attach candle*, to *support candle* concurrent with its generation of **s-c-b**. This interactive, simultaneous emergence of goal and new problem state is characteristic of AHA!'s behavior at the micro-level. More generally, such a simultaneous switch of two or more elements can be interpreted as a change in representation or restructuring. Depending on the values of the inhibition parameters used, AHA! can be seen to switch its representation from box-of-tacks to tacks and box at the RA level, simultaneously with the generation of the partial solution, **s-c-b** (support candle with box).

Conclusion

The themes of interactive activation and the interconnectedness of units are of the essence of each of the problem solving phenomena discussed above. Stress one part of the model by making a goal or problem feature more salient and the model shifts its behavior dynamically and interactively. This dynamic adaptation, so characteristic of human problem solvers, is reflectec both at the micro-level in the generation of a single state, and at the macro-level by the variety of interesting solution paths AHA! produces.

Acknowledgments

I would like to thank (and hold blameless) Jay McClelland, Mark St. John, and Herb Simon for their very useful comments and suggestions.

References

Glucksberg, S. & Weisberg, R. W. (1966) Verbal behavior and problem solving: some effects of labeling in a functional fixedness problem. *Journal of Experimental Psychology*, Vol. 71, No.5, 659-664.

Kaplan, C. A. (1986) *Cue driven shifts in representation.* Unpublished master's thesis. Carnegie-Mellon University.

McClelland, J. L. (1981) Retrieving general and specific information from stored knowledge of specifics. *Proceedings of the Third Annual Meeting of the Cognitive Science Society*, 170-172.

McClelland, J. L. & Rumelhart, D. E. (1988) *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises.* Cambridge, MA:MIT Press/Bradford Books.

Newell, A. & Simon, H. A., (1972) *Human Problem Solving.* Englewood Cliffs, NJ:Prentice-Hall.